13.6

Reprinted from the Preprint Volume of the Fifth Conference on Satellite Meteorology and Oceanography, September 3-7, 1990, London, England. Published by the American Meteorological Society, Boston, Mass., USA

A CASE STUDY ANALYSIS OF CLUSTERING APPLIED TO VAS MEASUREMENTS

Donald W. Hillger[1]
John S. Snook[2]
James F.W. Purdom[1]

[1]NESDIS/RAMM Branch, E/RA25
Cooperative Institute for Research in the Atmosphere (CIRA)
Colorado Sate University
Fort Collins, CO 80523-0033

[2]Forecast Systems Laboratory
National Oceanic and Atmospheric Administration
Boulder, CO 80303

1.0 INTRODUCTION

A joint effort is underway to test a technique for clustering satellite sounding measurements (Hillger and Purdom, 1990) that was developed at the Cooperative Institute for Research in the Atmosphere (CIRA) and applied to VISSR Atmospheric Sounder (VAS) data ingested at the Forecast Systems Laboratory (FSL). Some spatial averaging of VAS measurements is necessary to reduce random noise to specified sounding requirements for producing temperature and water vapor profiles. Clustering offers the advantage of increasing signal-to-noise by averaging of measurements which are similar to within the noise levels of the VAS instrument. At the same time, this averaging of similar measurements does not smear horizontal gradients in the data, thereby saving mesoscale information which might otherwise be destroyed by averaging in arbitrary field-of-view (FOV) blocks.

A case study day was chosen, which differs in two ways from previous tests of clustering on VAS. One difference was that measurements in selected VAS channels are not available for all horizontal scans of the satellite. As a result, not all VAS channels are available for each FOV. However, since clustering can be based on either all or a subset of the VAS channels, two options are available: 1) apply clustering to only those channels which are available at all FOVs; or, 2) apply clustering to only those FOVs where all channels are available. The first option was used in this study. The second difference was this case study included cloud-contaminated FOVs along with clear FOVs. Thus far, clustering has been tested for cases with cloud-contaminated FOVs removed from consideration. A goal of this study was to test clustering on a dataset which contained many cloud-contaminated FOVs. Clustering was used to group the cloud-contaminated FOVs in the same manner used to group clear FOVs. Cloud-contaminated VAS measurements could then be treated in groups, with the affected clusters either eliminated or treated as cloud-contaminated in the retrieval algorithm.

Findings from this study are the nearest to a real-time test that has been done. Results will define some of the advantages and disadvantages of clustering as compared to arbitrary blocking of VAS measurements which is presently used for operational satellite sounding production.

2.0 VISSR ATMOSPHERIC SOUNDER (VAS)

VISSR Atmospheric Sounder data consists of 12 infrared channels which respond to temperature and water vapor variations throughout the atmosphere. The infrared sensors come in two latitudinal FOV sizes, small (8 km at 40 degrees latitude) and large (16 km) resolution. This is coupled with a scan pattern which depends on the VAS channel resolution. Large resolution channels skip every other scan line when compared to small resolution channels. The strategy for small resolution channels is to scan four lines and skip four lines providing a 'venetian blinded' scene. During 1985, the 11.2 um window (band 8) channel was an exception in that all lines were scanned with a small resolution FOV. All channels have an 8 km FOV in the longitudinal direction. A given location will then contain either all or a subset of the VAS channels. The reason for mentioning this is related to the flexibility of the clustering technique described below.

2.1 Case Study Day (18 July 1985)

The case study day was 18 July 1985 when GOES-6 was at 98 degrees West longitude. The view from GOES-6 is shown in Figure 1, which is an 1100 UTC image from the VAS window channel (11.2 um). This case study was first treated by Snook and Birkenheuer (1986). The area of concern for this study was the Oklahoma and Texas panhandle region as shown in Figure 2. Shown are locations of VAS FOVs at 16 km resolution, the resolution of the large VAS sensors. Each FOV contains measurements from only 7 of the 12 VAS channels, due to the missing scan lines for some channels. It was decided to treat all the FOVs by ignoring some channels rather than to treat only those FOVs where all the channels are present. In this case there are 24 lines of 35 elements each, or 840 FOVs.

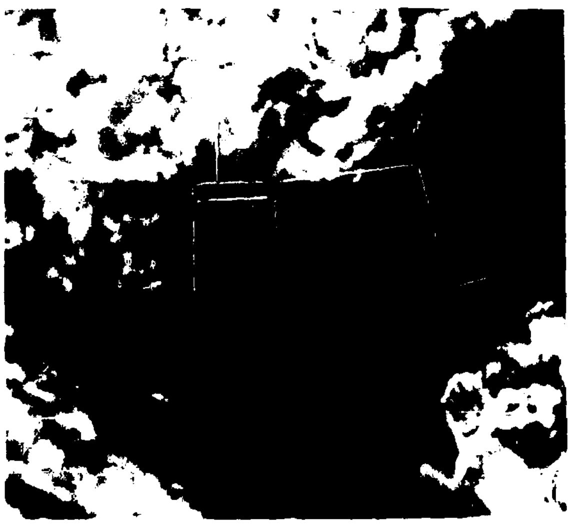

Figure 1: GOES window channel (11.2 um) image for 1100 UTC on 18 July 1985.

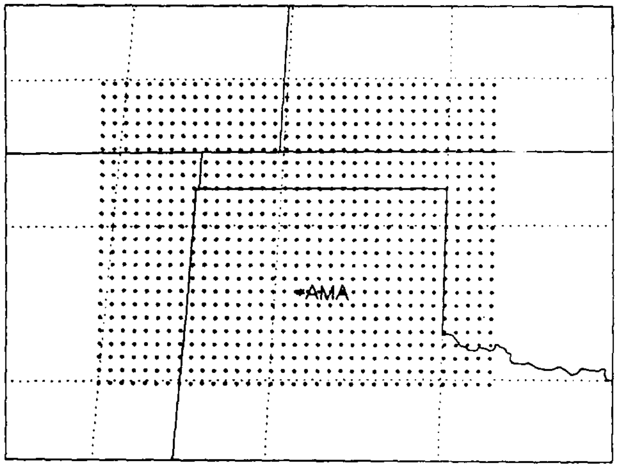

Figure 2: Area of concern showing VAS FOVs at 16 km resolution over the Oklahoma and Texas panhandle region. There are 24 lines of 35 elements at 16 km resolution, or 840 FOVs covering an area of about 500 km on a side. Also designated is the location of the Amarillo, Texas rawinsonde.

3.0 CLUSTER ANALYSIS

The cluster technique is described in Hillger and Purdom (1990). The method groups FOVs into clusters which are similar to within the noise levels of the channels being considered. The method used here is unlike other applications of clustering which group data based solely on similarity without regard to cluster size. In this case the cluster size is set by the noise levels of the channels. All measurements within a given cluster will then be similar to within the noise levels of the VAS measurements, and measurements in a given cluster will be different from those in another cluster by changes greater than the noise levels in the VAS measurements.

Clustering is a multivariate technique which considers more than one variable (or channel) in the grouping process. All or a subset of the VAS channels can be used in the grouping process. This flexibility allows clustering to be used in the case of VAS where not all the channels are available at every FOV. In this particular case, only the VAS channels which were available at all FOVs were used. As a result, VAS channels 3,4,5, and 7 were not used for clustering due to their venetian blinded scan pattern. However, these channels were used in the production of the VAS retrievals.

3.1 Principal components

Rather than clustering directly on the VAS channels, it is useful to first transform the VAS channels into VAS principal components. This reduces the number of variables which contain significant information (Hillger and Purdom, 1989a and 1989b). Because of the redundancy in VAS channel information, the 12 VAS channels can be transformed into about 5 principal components which contain 99% of the information content. In this case, VAS channels 1 and 2, which peak high in the upper atmosphere, were eliminated to reduce the dependence of the principal components on these noisy channels. The principal component transformation was then made from VAS channels 6,8,9,10, and 12. (Channel 11 data were missing due to instrumental problems.) The first three principal components now contain 99% of the information content, and the first two components alone contain about 80% of the information.

The first two principal components were used in this case for clustering. Figure 3a shows a scatter plot of the FOVs in principal component space. Small letters (A-Z) designate the cluster, with un-clustered FOVs designated by pluses (+). The clusters are easier to distinguish in Figure 3b where the ellipses represent the cluster extent, and the letter within each ellipse is the cluster identification. The 26 clusters are ordered alphabetically by the number of FOVs in each cluster, with the first cluster containing 77 FOVs, and the last cluster having only six FOVs. The two columns on the right side of Figure 5b give the number of FOVs in each cluster. Further clusters could have been chosen, but clusters with fewer FOVs are those which are less significant and which typically are cloud-contaminated.

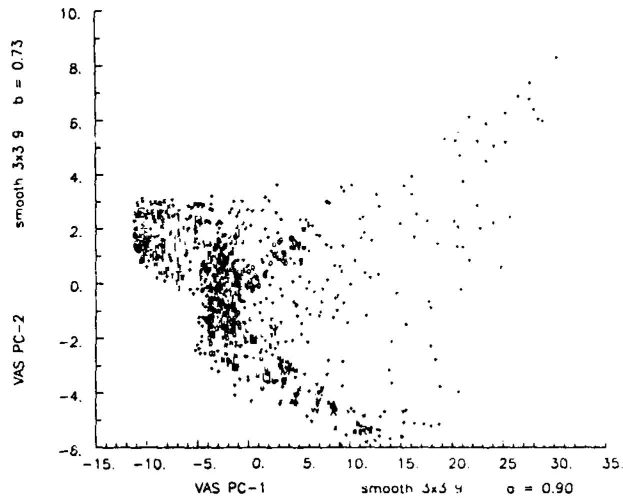

Figure 3a: Scatter diagram of VAS FOVs in principal component space. Clustered FOVs are designated by letters (A-Z), with un-clustered FOVs designated by pluses (+).

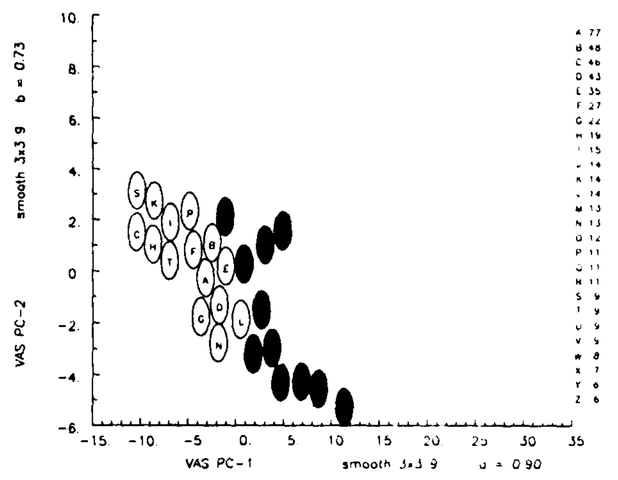

Figure 3b: Cluster extent represented by ellipses for the same area in principal component space as in Figure 3a. Shaded clusters are those in which FOVs have been determined to be cloud contaminated. The two columns on the right side give the number of FOVs in each cluster.

3.2 Cloud Clearing

Clouds have always been a problem for infrared sounding measurements from space. Unless the cloud-contaminated FOVs are identified, the retrieval algorithm can produce erroneous results. However, the clustering process offers hope for cloud clearing by treating similar measurements as a group. A retrieval can be generated for possibly cloudy measurements and the results can either be retained or eliminated for the whole cluster depending on the retrieval outcome.

Alternatively, cloud clearing can be based on the VAS measurements (effective blackbody temperatures) directly. That together with eliminating wayward retrievals resulted in designating some of the clusters as cloud-contaminated. Cloud clearing can be used more effectively on clustered FOVs, since the cluster mean has characteristics which are reinforced by the large number of FOVs in each cluster (typically 10 or more FOVs). In particular, low values in the window channels (VAS 8 or 12) are used for cloud clearing. The cloud contaminated clusters are shaded in Figure 3b. It is interesting that the cloud-contaminated clusters are separated on one side from the clear clusters. This makes sense, since the cloud contamination reduces the VAS effective blackbody temperatures, which in this case corresponds to a larger value for the first principal component.

The same clusters as in Figure 3 are shown in Figure 4 in line-element space corresponding to the FOV locations in Figure 2. Shading is again used to designate the clustered FOVs which were determined to be cloud-contaminated. Not all of the un-clustered FOVs, designated by pluses, are cloud-contaminated. Some pluses designate FOVs which fall between clusters and are in groups too small to form another cluster. Most of the cloudiness on the northern edge is easily seen in the window channel image in Figure 1, but some of the smaller shaded patches may be due to cirrus. This figure shows that the clusters with larger numbers of FOVs are clear.

After clustering is complete, one set of VAS measurements is used to represent each cluster. The set of measurements consists of the average values in each of the VAS channels, independent of whether that channel was used in the cluster analysis. Remember that some of the channels were not used because they were not available at all FOVs. However, each cluster does contain FOVs with all VAS channels.

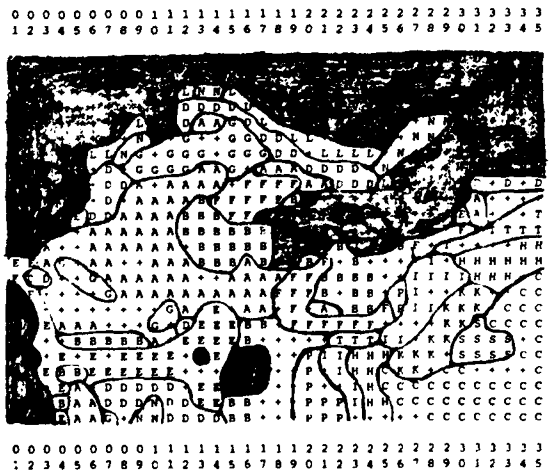

Figure 4: The same clustered FOVs as in Figure 3 but in line-element space. Shading is again used to designate cloud-contaminated FOVs.

4.0 FSL RETRIEVALS

FSL generates operational satellite soundings in real time using the physical simultaneous retrieval algorithm (Hayden, 1988) developed at the Cooperative Institute for Meterological Satellite Studies (CIMSS). Retrievals are located within swaths of satellite data at a 56 km resolution (Snook, 1989). Surface and upper air first guess information is provided by FSL's Mesoscale Analysis and Prediction System (Benjamin, et al., 1990) which incorporates wind profiler and aircraft reports in addition to conventional data. Objective cloud clearing is accomplished through a comparison of the 11.2 um window (VAS channel 8) effective blackbody temperature with the first guess surface temperature (Snook, 1987) for each satellite FOV. If the satellite effective blackbody temperature is more than ten degrees Celsius colder than the surface temperature, that satellite FOV is not used in the block averaging. At least 33% of the satellite FOVs within the averaging block must be clear for a retrieval to be produced at that location.

5.0 CLUSTERING VERSUS BLOCKING

The intent of clustering is to reinforce information within the VAS FOVs by grouping together similar measurements, while at the same time reduce the amount of smoothing through existing gradient information. Therefore, clustering is an improvement to present operational techniques which use an arbitrary rectangular block of FOVs to produce one retrieval. The need for several FOVs is to reduce noise by averaging together VAS measurements, and to look for cloud-contaminated FOVs. The operational retrievals generated for this study used blocks of FOVs consisting of 3 lines of 6 elements, or 18 FOVs at 16 km resolution. This compares to an average of about 25 FOVs for clear clusters. Thus, the clusters contain more FOVs than the blocks, which is beneficial for reduced noise. Furthermore, the clusters do not average together measurements which vary widely. Rather the measurements are all within the noise level of each other. The same cannot be said for the blocks of FOVs which are currently used operationally.

The same retrieval scheme was used to produce soundings from both the clustered and the blocked VAS measurements. The only difference was that in one case clustering was used to group the VAS measurements, and in the other case the VAS measurements were blocked into arbitrary groups of FOVs as are used operationally.

5.1 Individual retrievals

The only rawinsonde within the analysis area is Amarillo, Texas (AMA) taken at 1200 UTC, within about an hour of the time of the VAS measurements. Two skew-T, log-P plots are used to compare the AMA sounding to the VAS retrievals. In Figure 5a, the comparison is between the AMA sounding (solid) and the retrieval produced from VAS measurements (dashed) for the nearest cluster (cluster F). In Figure 5b the comparison is between the AMA sounding (solid) and the retrieval produced from VAS measurements (dashed) for the nearest rectangular block. Both satellite retrievals do a respectable job of reproducing the temperature structure around AMA, with the clustered retrieval being closer at the higher levels. In both cases, the AMA dewpoint temperatures are not faithfully reproduced, but the cluster retrieval seems slightly better in the lower levels. Remember that both satellite retrievals use sets of VAS measurements. The only difference is the grouping of the measurements by either clustering or blocking.

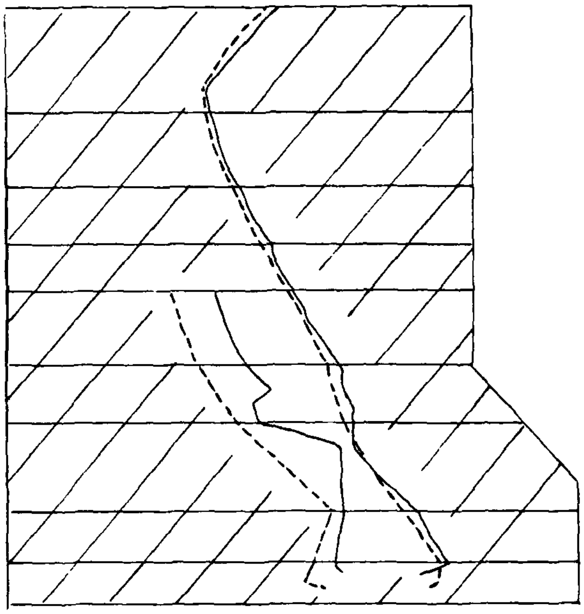

Figure 5a: Comparison of the Amarillo, Texas sounding (solid) at 1200 UTC to a VAS retrieval (dashed) for the nearest cluster (cluster F).

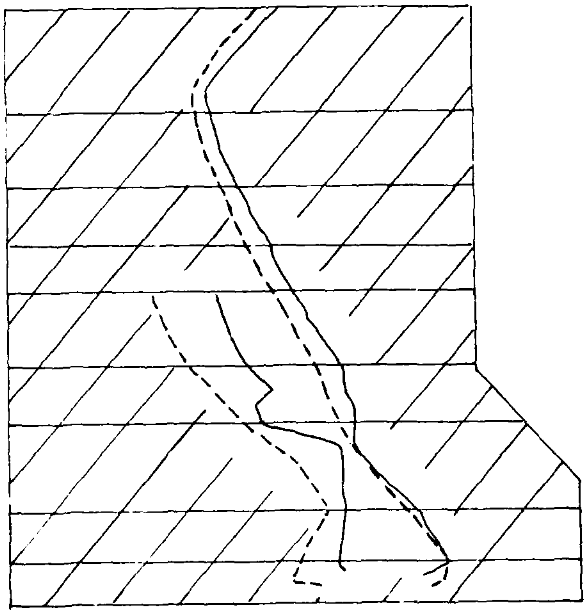

Figure 5b: Same as Figure 5a, but for the VAS retrieval for the nearest block.

5.2 Horizontal fields

Once the retrievals are generated, the retrieved soundings for each cluster are used to reconstruct the entire field. This is accomplished by considering the cluster to which each FOV belongs, as well as adjacent clusters. A special interpolation scheme is used to determine the value at each FOV based on its distance in cluster space to the nearest three cluster centers (Hillger and Purdom, 1990). An example of a field produced from retrievals on clustered FOVs is shown in Figure 6a, which is the 850 hPa temperature analysis over the analysis area. The equivalent 850 hPa temperature analysis produced from retrievals on blocked FOVs is shown in Figure 6b. Significant differences between the two analyses can be seen. In particular, the blocked retrievals show more small-scale variability, possibly due to cloud contamination. It appears that cloudy FOVs have been more successfully eliminated with the clustering process. However, the same general gradient exists in both figures, with warmer temperatures in the south and east.

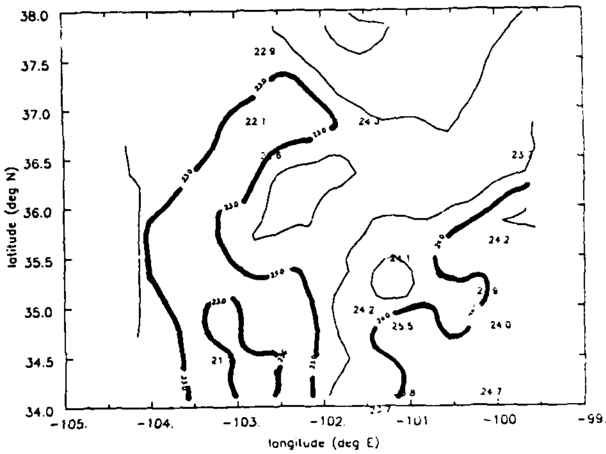

Figure 6a: Field of 850 hPa temperatures produced from retrievals on clustered VAS measurements. Contours are every 1 degree Celsius.

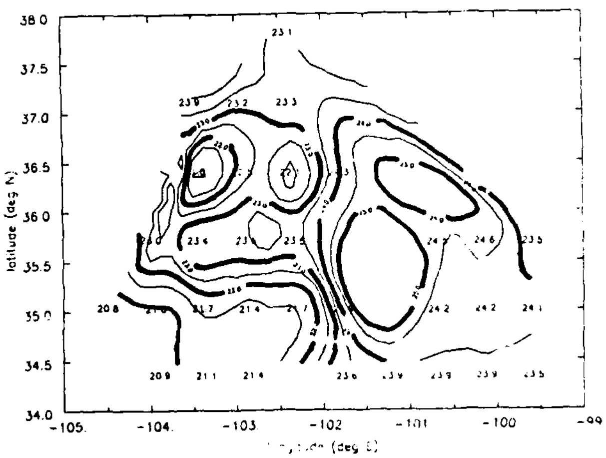

Figure 6b: Same as Figure 6a, but from retrievals on blocked VAS measurements.

Differences between clustered and blocked retrievals are shown in Figure 7 (a and b) for 850 hPa dewpoint temperatures. Here both figures show more moisture to the northeast, but the blocked retrievals show extreme drying to the west, which is not shown by the clustered retrievals. Again, this may be due to cloud contaminated blocks not being eliminated.

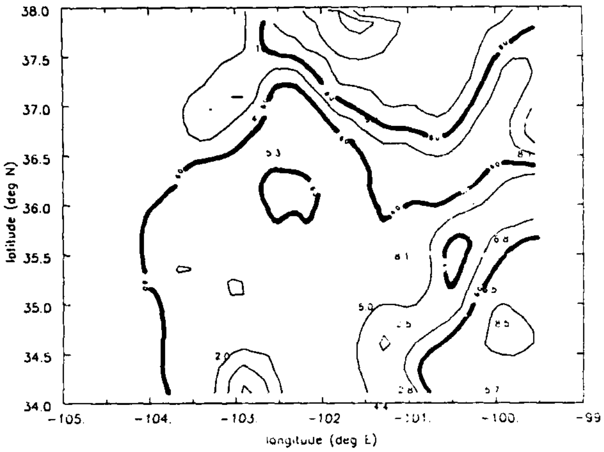

Figure 7a: Same as Figure 5a, but for 850 hPa dewpoint temperatures. Contours are every 2 degrees Celsius.

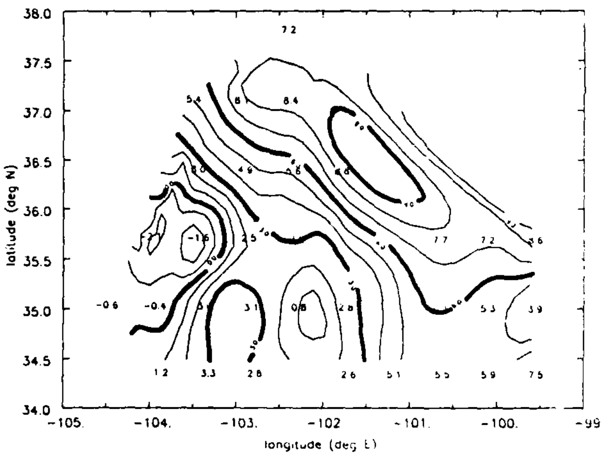

Figure 7b: Same as Figure 7a, but from retrievals on blocked VAS measurements.

Figure 8 (a and b) shows the 500 hPa temperature analyses for the clustered and the blocked retrievals. As was the case with the 850 hPa temperatures, the blocked retrievals show more small scale variability which again may be due to cloud contamination. However, the larger-scale features are vaguely similar.

Finally, in order to compare the two sets of satellite retrievals, the total-totals stability index (degrees Celsius) was computed for each of the fields. A difference plot is shown in Figure 9. This is the difference in stability (clustered minus blocked retrievals), showing a large gradient in stability in the difference field. This large difference indicates that the way VAS data is handled strongly affects the results of the retrievals.

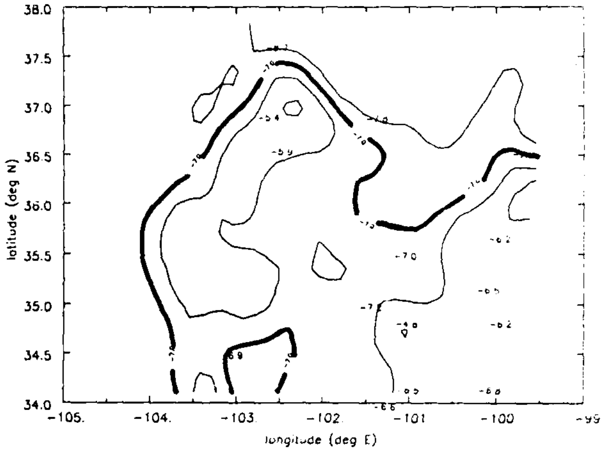

Figure 8a: Same as Figure 6a, but for 500 hPa temperatures. Contours are every 1 degree Celsius.

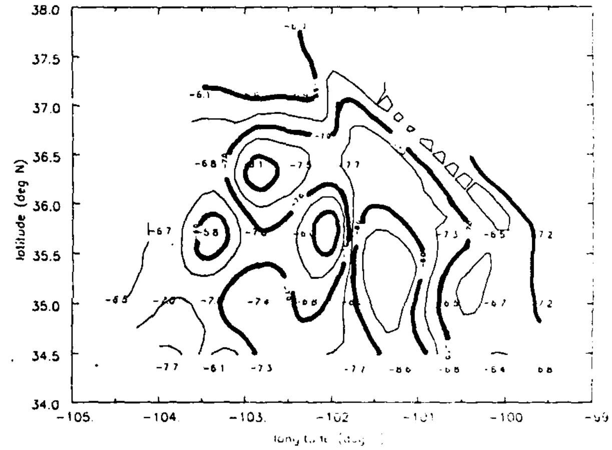

Figure 8b: Same as Figure 8a, but from retrievals on blocked VAS measurements.

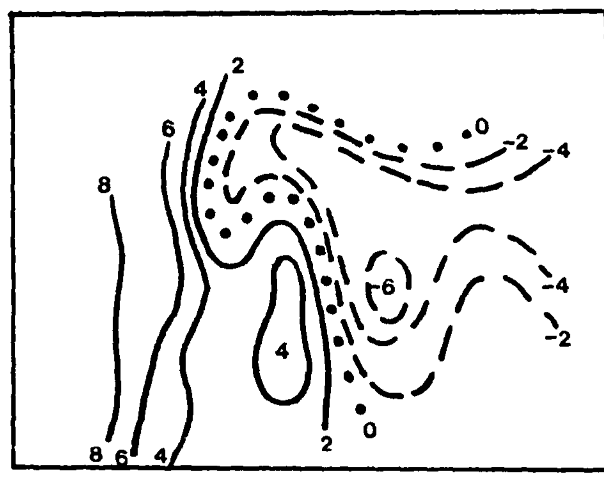

Figure 9: Field of total-totals stability index differences between retrievals on VAS measurements (clustered minus blocked). Contours are every 2 degrees Celsius.

VISSR Atmospheric Sounder

6.0 SUMMARY AND CONCLUSIONS

This study is a joint effort to test clustering of VAS measurements in a more operational setting. The operational setting differs from previous tests of clustering. The clustering technique had to handle VAS measurements which were not spatially continuous, due to operational scanning requirements for skipping certain scan lines. This was handled by considering only those VAS channels which were available at every FOV. The operational test also dealt with cloud-contaminated FOVs. Clustering can detect cloud-contaminated VAS measurements by treating them as a group and eliminating clusters which are either outliers or which produce suspect soundings.

A comparison was made between retrievals produced from clustered VAS measurements to retrievals produced from blocked VAS measurements. Differences between retrievals using the two methods are significant, especially considering the small area of concern. Neither set of retrievals can necessarily be shown to be better, due to a lack of conventional measurements for comparison at such high resolution. However, slight improvements in retrievals can be expected due to increased signal-to-noise of the VAS measurement which are clustered as compared to present blocking schemes. Further testing will be performed on this and other data sets.

(RH)

ACKNOWLEDGEMENTS

Invaluable computing assistance was provided by Debra Lubich of CIRA. This work was supported by NOAA Grant NA85RAH05045.

REFERENCES

Benjamin, S.G., K.A.Brewster, R.Brummer, B.F.Jewett, T.W.Schlatter, T.L.Smith, and P.A.Stamus, 1990: A real-time 3-hourly isentropic mesoscale data-assimilation system using ACARS aircraft data combined with other observations. Submitted to Mon. Wea. Rev.

Hayden, C.H., 1988: GOES-VAS simultaneous temperature-moisture retrieval algorithm. J. Appl. Meteor., 27, 705-733.

Hillger, D.W., and J.F.W.Purdom, 1989a: Clustering of satellite sounding radiances to maximize mesoscale meteorological detail. Fourth Conference on Satellite Meteorology and Oceanography, AMS, 16-19 May, San Diego, CA, 228-231.

Hillger, D.W., and J.F.W.Purdom, 1989b: Using VAS data to add mesoscale detail to synoptic-scale radiosonde data. Twelfth Conference on Weather Analysis and Forecasting, AMS, 2-6 October, Monterey, CA, 412-415.

Hillger, D.W., and J.F.W.Purdom, 1990: Clustering of satellite sounding radiances to enhance mesoscale meteorological retrievals. J. Appl. Meteor., 29, 13 p. [accepted for publication]

Snook, J.S., and D.Birkenheuer, 1986: Utilization of VAS data in PROFS 1985 forecasting exercise. *Eleventh Conference on Weather Forecasting and Analysis*, AMS, 17-20 June, Kansas City, MO, 155-159.

Snook, J.S., 1987: Operational cloud-clearing techniques of VAS radiance data. Symposium on Mesoscale Analysis and Forecasting, ESA SP-282, IUGG, 17-19 August, Vancouver, Canada, 103-106.

Snook, J.S., 1989: An evaluation of operational VAS soundings at *PROFS and their application on the* local scale. Twelfth Conference on Weather Forecasting and Analysis, AMS, 2-6 October, Monterey, CA, 500-504.